CHAPTER 6

How stable are extreme P values: an empirical illustration

André R. de Vries[1], Robert M.W. Hofstra[1], Lude Franke[1], Gerard J. te Meerman[1]

[1] Department of Genetics, University Medical Center Groningen and University of Groningen, the Netherlands

Submitted for publication

Abstract

Whole genome association studies focus on the lowest P values found for hundreds of thousands of markers. Using a dataset of 771 case and 1,417 control genotypes on celiac disease, we assessed the effects of choosing different subsets of 771 out of the 1,417 controls on the association test outcomes. We observed a large variation of resulting P values (10^{-2} to 10^{-9}) with a specific SNP depending on the choice of control subset. Our results indicate that extreme single marker association P values in studies might have low reproducibility for typical medium sized control populations. Likewise, case populations are expected to show similar effects. P value variation appeared to be independent of allele frequency. Larger study sizes are required in order to improve the reliability of P values, but large studies encounter potential loss of power due to genetic heterogeneity.

Introduction

Whole genome association studies focus on the lowest P values found for hundreds of thousands of markers. Because of 'between SNP correlations' it is difficult to assess the required multiple testing correction needed to avoid false positive results. Ioannidis has consistently shown theoretically that most reported association results are untrue or inflated [1,2]. We show here empirical results for the stability of P values by resampling from a control dataset to assist in interpreting P values.

Methods

Recently, a number of new genes have been found to be associated with Celiac Disease (CD), using 778 selected cases and 1,422 controls from a control cohort [3]. In a second cohort these data were replicated [4]. We reanalyzed these data in order to assess the variability of P values in different control sample subsets. For our study, we rejected 7 cases and 5 controls that showed call rates below 98%. We selected all 771 cases and 10,000 different subsets of 771 controls by resampling with replacement. P values were calculated using the allelic χ^2-test.

Results

We plotted the distribution of -log P values of three SNPs of interest (all in the IL21 region) resulting from using 10,000 different control subsets of 771 controls (figure 1). A wide distribution of -log P values was found, depending on the choice of control subset. This choice determines whether a SNP will show a significant result or not.

In χ^2-testing with low numbers in one of the cells, adding or removing a few people can have dramatic effects on P values. Therefore, low frequent SNPs are expected to show higher variability. Yet, we found that this is not the case, as shown in figure 2. We observed no correlation of the minor allele frequency of 200 SNPs with the variance of the resulting P value distribution from 1,000 tests at each SNP (Pearson $r^2 = 0.00029$, P value for correlation = 0.41).

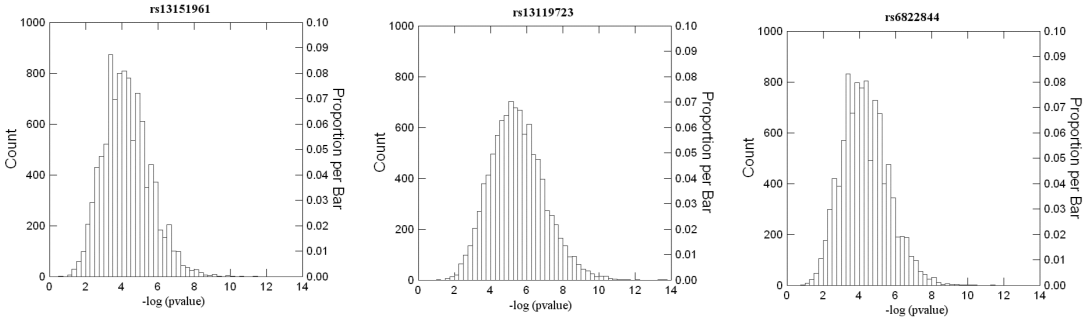

Figure 1: *Distribution of -log P values for the three SNPs of interest resulting from 10,000 random selections using replacement of 771 control samples. Standard deviation of the P value distribution for the three SNPs are 1.28, 1.48 and 1.28 respectively.*

Discussion & conclusion

Our results show that large variation of P values is observed for specific SNPs, depending on the sampling of subjects. There are many SNPs in a genome wide association test, and by just adding or removing a single control genotype, many SNPs will have their allele frequency slightly altered in the control group, leading to different P values. So, the choice of controls, but the same holds for patients, affects resulting P values. Especially smaller studies (<1,000 samples) are likely to be based on sample sets that are not a good reflection of the larger population for at least some SNPs. In other words: the selection of a study

cohort (which is a subset of the total population) could for some SNPs result in P values in the right hand extreme of the distribution. Those SNPs may falsely be called associated with the phenotype.

It is, therefore, better to increase sample sizes in order to minimize the variance of P values. This will increase the chance that the studied cohort reflects the true population and increases the chance that the "sampled" P values are close to the true values. However, although the use of large control cohorts, for example from central databanks, can increase the control group enormously and make P values more reliable, the involved introduction of more genetic heterogeneity also reduces power. In addition, genetic heterogeneity can affect P value variation as well [5-7]. Therefore, care must be taken in order to maximize the size of the dataset while introducing as little genetic heterogeneity as possible.

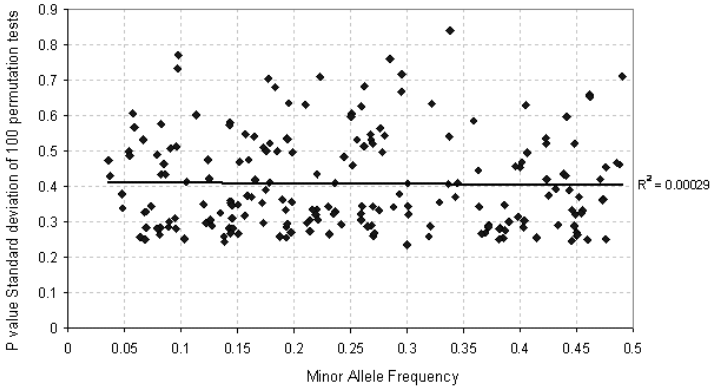

Figure 2: *P value standard deviation of 1,000 random control selection tests plotted as a function of the minor allele frequency for 200 SNPs. Each dot represents one SNP. Correlation coefficient r = 0.017 and the P value for directional correlation is 0.41.*

References

1. Ioannidis, J. P., 2005 **Why most published research findings are false**. *PLoS.Med.* 2: e124.
2. Ioannidis, J. P., 2008 **Why most discovered true associations are inflated.** *Epidemiology* 19: 640-648.
3. van Heel, D. A., L. Franke, K. A. Hunt, R. Gwilliam, A. Zhernakova et al. 2007 **A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21.** *Nat.Genet.* 39: 827-829.
4. Hunt, K. A., A. Zhernakova, G. Turner, G. A. Heap, L. Franke et al. 2008 **Newly identified genetic**

risk variants for celiac disease related to the immune response. *Nat.Genet.* 40: 395-402.

5. Ioannidis, J. P., N. A. Patsopoulos, and E. Evangelou, 2007 **Heterogeneity in meta-analyses of genome-wide association investigations**. *PLoS ONE*. 2: e841.
6. Moonesinghe, R., M. J. Khoury, T. Liu, and J. P. Ioannidis, 2008 **Required sample size and nonreplicability thresholds for heterogeneous genetic associations**. *Proc.Natl.Acad.Sci.U.S.A* 105: 617-622.
7. Patsopoulos, N. A., E. Evangelou, and J. P. Ioannidis, 2008 **Sensitivity of between-study heterogeneity in meta-analysis: proposed metrics and empirical evaluation**. *Int.J.Epidemiol.* 37: 1148-1157.